NAWCWD TP 8349
NAVAIR

Multichannel Graduated Non-Convexity Theory Summary

by
Jorge M. Martin
Research Department

SEPTEMBER 2002

NAVAL AIR WARFARE CENTER WEAPONS DIVISION
CHINA LAKE, CA 93555-6100

Naval Air Warfare Center Weapons Division

FOREWORD

This report documents the mathematical foundation needed to extend to the multichannel setting in dimensions 1 and 2 a method for segmentation, boundary detection, and smoothing known as the Graduated Non-Convexity Method. The work was done at the Naval Air Warfare Center Weapons Division, China Lake, California, from October 1995 to January 1996. It was funded by the NAWC In-House Laboratory Independent Research (ILIR) Program under Program Element 601152N, Work Unit Numbers 12304016 and A47D47, ILIR Project Multi-scale Segmentation Techniques.

Approved by
C. F. MARKARIAN, *Head*
Research Department
1 September 2002

Under authority of
MICHAEL C. BACHMANN
RDML, U.S. Navy
Commander

Released for publication by
KAREN HIGGINS
Director for Research and Engineering

NAWCWD Technical Publication 8349

Published by . . . . .Technical Information Division
Collation . . . . Cover, 21 leaves
First printing . . . . .29 copies

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
September 2002

3. REPORT TYPE AND DATES COVERED
October 1995 to January 1996

4. TITLE AND SUBTITLE
Multichannel Graduated Non-Convexity Theory Summary (U)

5. FUNDING NUMBERS
PE 601152N
WM 12304016
WM A47D47

6. AUTHOR(S)
Jorge M. Martin

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Air Warfare Center Weapons Division
China Lake, CA 93555-6100

8. PERFORMING ORGANIZATION REPORT NUMBER
NAWCWD TP 8349

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12A. DISTRIBUTION/AVAILABILITY STATEMENT
A Statement; public release; distribution is unlimited.

12B. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

(U) This paper contains the mathematical foundation needed to extend the graduated non-convexity (GNC) method to the multichannel setting (vector valued data) in dimensions 1 and 2. The theory is presented as a collection of definitions and propositions with proofs. The result for dimensions 3 or greater is stated without proof; the proof follows the same steps as in dimensions 1 and 2. The most important technical issue to be resolved is whether the convex approximation exists and how to obtain it. An outline of the GNC method in the one dimensional, single channel setting is included for completeness.

14. SUBJECT TERMS
Boundary Detection; Multichannel; Multi-Scale; Segmentation; Variational Methods; Weak Continuity Constraint

15. NUMBER OF PAGES
42

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	U L

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

CONTENTS

1. INTRODUCTION

Segmentation and boundary detection algorithms are basic tools for extracting global features out of digitized data. Techniques based on variational methods achieve the result (a segmented image) by minimizing a cost functional. In 1985, Mumford and Shah introduced a mathematical model that captures the essential features that must be considered when solving the segmentation problem (Reference 1). It is a *multi-scale* technique that allows the extraction of features at different levels of detail (scale). It is also a *multichannel* technique that can be used to segment images of a scene when registered multiple data channels for the same scene are available. These may be data channels from various sensors, hue channels, preprocessing channels such as wavelet or other transform channels, etc.

The *Mumford & Shah* (*M&S*) *Functional* (References 1 and 2) has the form

$$E(u,K) \equiv \alpha^2 \int_{\Omega - K} \|u-g\|^2 d\mu + \int_{\Omega - K} \|\nabla u\|^2 d\mu + \lambda \cdot \ell(K), \tag{1.1}$$

where Ω denotes a "rectangle" in R^d, $d \in \{1,2,3,\cdots\}$, $g:\Omega \to R^c$ denotes the "image" (g is μ-measurable), c = number of "channels", $c \in \{1,2,3,\cdots\}$, and $u:\Omega \to R^c$ is an approximation to g belonging to a set of functions Φ. The rectangle Ω is decomposed into a finite collection of disjoint open sets O_n (n = 1, 2, ..., N) and their boundaries $K = \bigcup_{n=1}^{N} \partial O_n$; so that $\Omega = \left[\bigcup_{n=1}^{N} O_n\right] \cup K$, and K is sufficiently nice to have a "length", denoted by $\ell(K)$. The disjoint open sets O_n (n = 1, 2, ..., N) and their boundaries $K = \bigcup_{n=1}^{N} \partial O_n$ together with u are called a *segmentation* of (Ω, g).

The goal is to construct 2 things:

(a) A smoothed ideal image, $u : \Omega \to R^c$.
(b) A set of boundaries, $K \subset \Omega$;

u and K are found by minimizing the functional E.

The first term on the right hand side (RHS) of Equation 1.1 ensures that u is a faithful representation of g, the second term ensures that u is as smooth as possible on each open set O_n, and the last term prevents the boundaries from growing too large. The parameters α and λ are weighting factors that control the quality of the approximation and the coarseness of the segmentation.

The *Weak Continuity Constraint* models of Blake and Zisserman are another collection of effective techniques for segmentation, boundary detection, and smoothing of data (Reference 3). In particular, the *weak string constraint* and the *weak membrane constraint* can be considered to be the discrete forms of the M&S functional in dimensions 1 and 2, respectively. Blake and Zisserman developed a technique called *Graduated Non-Convexity (GNC)* (Reference 3) that can be used to minimize the discrete M&S functional.

The GNC method relies on a homotopy between the cost functional and a convex approximation to it. The minimum of the convex approximation is easily found. The convex functional is gradually deformed via the homotopy until it converges to the original functional. At each step in the gradual deformation a minimization is performed. The collection of minimal points so obtained forms a path that leads to a minimum of the original functional. However, this technique was developed only for single channel data.

In this paper we extend the GNC method to the *multichannel setting* (that is, *vector valued data*). We develop the mathematical foundation for the technique in the multichannel setting. The most important technical issue of this method is the convex approximation: whether it exists and how it is obtained. In the multichannel setting, in dimensions 1, 2, and higher, determining if the convex approximation exists, as well as the specific details of how to obtain it, are the problems to be resolved.

A summary of the theory behind the GNC method in the multichannel setting for dimensions 1 and 2 will be presented here as a collection of definitions and propositions with proofs. The results for dimension $d \geq 3$ will be stated here without a proof. The proof follows the same steps as the proofs for $d = 1, 2$.

Section 2 contains an outline of the GNC method. Section 3 contains a complete summary of the multichannel theory for dimension 1 in the form of a series of propositions that lead to the convexity result, *Proposition 3.8*. In Section 4 we present the theory for dimension 2 in the same form, leading to *Proposition 4.4*.

2. THE GRADUATED NON-CONVEXITY METHOD

For completeness and to introduce the reader to the GNC method, the idea behind the method is briefly described in this section in the one-dimensional, single channel setting (see Reference 3 for more details).

The *Discrete Mumford & Shah Functional* for *one-dimensional* and *single channel* data $x = [x_1, x_2, \ldots, x_N]^T \in R^N$ has the form

$$E(u,a) = \sum_{i=1}^{N} (u_i - x_i)^2 + \lambda^2 \sum_{i=1}^{N-1} [u_{i+1} - u_i]^2 (1 - a_i) + \alpha \sum_{i=1}^{N-1} a_i \,, \tag{2.1}$$

where $u = [u_1, u_2, \ldots, u_N]^T \in R^N$, $a = [a_1, a_2, \ldots, a_N]^T$, $a_i \in \{0,1\}$ $\forall i$, and λ and α are positive weights. This is called *The Weak String Constraint* in Reference 3.

The three terms on the RHS of Equation 2.1 play, respectively, the same roles as the three terms on the RHS of Equation 1.1. Here, x plays the role of g.

We seek a vector u that is smooth and is close to x. The vector u is found by minimizing Equation 2.1 with respect to u and the Boolean variable a.

The minimization with respect to a can be done analytically (Reference 3), as follows.

Elimination of the Boolean Variables

Let $h: R \times \{0,1\} \to R$ be given by $h(t,b) \equiv \lambda^2 t^2 (1-b) + \alpha b \quad (t \in R, b \in \{0,1\})$, and let $g: R \to R$ be given by

$$g(t) \equiv \begin{cases} \lambda^2 t^2 & \text{if } |t| \le \sqrt{\alpha}/\lambda \\ \alpha & \text{if } |t| \ge \sqrt{\alpha}/\lambda . \end{cases}$$

Note that, by putting the last two summations in one, E can be written in terms of h as:

$$E(u,a) = \sum_{i=1}^{N} (u_i - x_i)^2 + \sum_{i=1}^{N-1} h(u_{i+1} - u_i, a_i) . \tag{2.2}$$

Moreover, since $g(t) = \min\limits_{b \in \{0,1\}} h(t,b) \quad \forall t \in R$, it follows that, for each $u \in R^N$, the minimum of E over a is given by

$$F(u) = \sum_{i=1}^{N} (u_i - x_i)^2 + \sum_{i=1}^{N-1} g(u_{i+1} - u_i), \quad u = [u_1, u_2, \ldots, u_N]^T \in R^N . \tag{2.3}$$

Theorem 2.1 $\quad \min\limits_{a \in \{0,1\}^N} E(u,a) = F(u) \qquad \forall u \in R^N .$

The Convex Approximation F_{λ_0}

The function F is non-convex. As noted above, the GNC method consists of approximating F with a convex function F_{λ_0} and obtaining a homotopy between F_{λ_0} and F. The minimum of the convex approximation F_{λ_0} is easily found (e.g., by steepest descent). The convex function is gradually deformed via the homotopy until it converges to the original function. At each step in the gradual deformation, a minimization is performed; the starting point in the minimization procedure is the minimum of the previous function. The collection of minimal points so obtained defines a path that leads to a minimum of F.

The homotopy is defined in terms of a collection of functions F_γ $(0 < \gamma_0 \le \gamma < \infty)$, such that $F_\gamma \to F$ uniformly as $\gamma \to \infty$.

The functions F_γ are given by (2.3) with g replaced by $g_\gamma : R \to R$ defined by

$$g_\gamma(t) \equiv \begin{cases} \lambda^2 t^2 & \text{if } |t| \le q \\ \alpha - \frac{1}{2}\gamma(|t| - r)^2 & \text{if } q \le |t| \le r \\ \alpha & \text{if } \mathrm{r} \le |t| \end{cases} \quad (\gamma > 0), \text{ with } q = \frac{\alpha}{\lambda^2 r} \text{ and } r^2 = \alpha\left(\frac{2}{\gamma} + \frac{1}{\lambda^2}\right).$$

Remark 2.1. $g_\gamma \to g$ uniformly as $\gamma \to \infty$.

The GNC theory consists of showing that there exists a positive value of γ ($\gamma = \gamma_0 = 1/2$) for which F_γ is convex. By letting $\gamma(t) = \frac{\gamma_0}{t}$ $(0 < \mathrm{t} \le 1)$, one obtains the desired homotopy $\mathrm{H}: R^N \times [0,1] \to [0,\infty)$.

$$\mathrm{H}(u,t) = \begin{cases} F(u) & \text{for} \quad t = 0, u \in R^N \\ F_{\gamma(t)}(u) & \text{for} \quad t \in (0,1], u \in R^N. \end{cases}$$

Remark 2.2. Since $F_\gamma \to F$ uniformly as $\gamma \to \infty$ and $F_\gamma(u)$ varies continuously with γ and u, the function H is continuous on $R^N \times [0,1]$ and, therefore, it defines a homotopy between F and F_{λ_0}.

In this paper we present the generalization of this theory to the multichannel setting in dimensions 1 and 2. The extension to dimensions 3 and higher follows the development of the theory in dimension 2 in Section 4. The value of γ for which F_γ is convex satisfies $\gamma = \frac{1}{2d}$, where $d = 1, 2, 3, \ldots$ is the dimension of Ω.

In the multichannel setting, each x_i and each u_i are vectors of dimension *c*, where *c* is the number of channels. The obvious thing to do is to introduce a vector norm (Euclidean Norm) in Equations 2.1 and 2.3, as in the M&S functional. This leads to the one-dimensional *c*-channel model

$$E(u,a) = \sum_{i=1}^{N} \|u_i - x_i\|^2 + \sum_{i=1}^{N-1} \left\{ \lambda^2 \|u_{i+1} - u_i\|^2 (1 - a_i) + \alpha a_i \right\}$$

and

$$F(u) = \sum_{i=1}^{N} \|u_i - x_i\|^2 + \sum_{i=1}^{N-1} g(\|u_{i+1} - u_i\|),$$

with $u = [u_1{}^T, u_2{}^T, \ldots, u_N{}^T]^T \in R^{Nc}$, $u_k \equiv [u_{k,1}, u_{k,2}, \cdots, u_{k,c}]^T \in R^c$ $(1 \le k \le N)$, and $a \equiv [a_1, a_2, \cdots, a_N]^T \in \{0,1\}^N$.

The theory follows the same steps outlined above. The greatest challenge was to establish convexity in the presence of the vector norm instead of a scalar absolute value. The theory does go through even in the presence of a weighted vector norm. It becomes evident that the value of γ for which F_γ is convex must satisfy

$\frac{1}{\gamma} = 2 \cdot (\textit{number of terms in } G)$ and that the dimension $d = (\textit{number of terms in } G)$. Thus, in general, $\gamma_0 = \dfrac{1}{2d}$, where d = 1, 2, 3, ... is the dimension of the domain Ω. (G is defined in Definition 3.10, with $\ell = N$, for $d = 1$ and in Definition 4.5 for $d = 2$.)

We now present a collection of definitions that constitute the generalization of the GNC method to the multichannel setting in dimension $d = 1$.

3. THE 1D-MULTICHANNEL GNC CASE

3.1. DEFINITIONS

3.1. $g : R \to R$ is given by $g(t) \equiv \begin{cases} \lambda^2 t^2 & \text{if } |t| \le \sqrt{\alpha}/\lambda \\ \alpha & \text{if } |t| \ge \sqrt{\alpha}/\lambda \end{cases}$, for $\alpha > 0$.

3.2. $g_\gamma : R \to R$ is given by $g_\gamma(t) \equiv \begin{cases} \lambda^2 t^2 & \text{if } |t| \le q \\ \alpha - \frac{1}{2}\gamma(|t| - r)^2 & \text{if } q \le |t| \le r \\ \alpha & \text{if r} \le |t| \end{cases}$ ($\gamma > 0$),

with $q = \dfrac{\alpha}{\lambda^2 r}$ and $r^2 = \alpha(\dfrac{2}{\gamma} + \dfrac{1}{\lambda^2})$.

3.3. $g^+ : R \to R$ is given by $g^+(t) \equiv -\frac{1}{2}\gamma t^2$ $(t \in R)$.

3.4. $h : R \times \{0,1\} \to R$ is given by $h(t,b) \equiv \lambda^2 t^2 (1-b) + \alpha b$ $(t \in R, b \in \{0,1\})$.

3.5. c denotes the number of channels per pixel; N denotes the number of pixels.

3.6. $D : R^{Nc} \to R$ is given by $D(u) \equiv \sum_{k=1}^{N} \|u_k - x_k\|_W^2$,

where $u_k \equiv [u_{k,1}\ u_{k,2}\ \cdots\ u_{k,c}]^T \in R^c$ $(k = 1, 2, \cdots, N)$, $u \equiv [\,u_1{}^T \vdots\, u_2{}^T \vdots \cdots \vdots\, u_N{}^T\,]^T \in R^{Nc}$, $W = diag(w_1, w_2, \cdots, w_c)$ $(w_i > 0)$, and the weighted norm $\|\ \|_W : R^c \rightarrow [0, \infty)$ is defined as $\|y\|_W^2 \equiv \sum_{i=1}^{c} w_i y_i^2$.

3.7. $L_k : R^{Nc} \rightarrow R^c$ is defined as $L_k(u) \equiv u_{k+1} - u_k$ for $k = 1, 2, \cdots, N-1$ and for $k = N$, $L_N(u) \equiv u_1 - u_N$ $\quad (u \equiv [\,u_1{}^T \vdots\, u_2{}^T \vdots \cdots \vdots\, u_N{}^T\,]^T \in R^{Nc})$.

3.8. $\widetilde{G} : R^{Nc} \times \{0,1\}^N \rightarrow R$ is given by $\widetilde{G}(u, a) \equiv \sum_{k=1}^{\ell} h(\|L_k(u)\|_W, a_k)$, with $\ell \leq N$, u as in **3.6**, and $a \equiv [a_1\ a_2 \cdots a_N]^T \in \{0,1\}^N$.

3.9. $\widetilde{F} : R^{Nc} \times \{0,1\}^N \rightarrow R$ is given by $\widetilde{F} \equiv D + \widetilde{G}$.

3.10. $G : R^{Nc} \rightarrow R$ is given by $G(u) \equiv \sum_{k=1}^{\ell} g(\|L_k(u)\|_W)$ $\quad (u \in R^{Nc})$, $\ell \leq N$.

3.11. G_γ and $G^+ : R^{Nc} \rightarrow R$ are defined as G with g replaced by g_γ and g^+, respectively.

3.12. F, F_γ, and $F^+ : R^{Nc} \rightarrow R$ are defined, respectively, as $F = D + G$, $F_\gamma = D + G_\gamma$, and $F^+ = D + G^+$.

3.2. CONVEXITY OF $F_\gamma : R^{Nc} \rightarrow R$

Proposition 3.1. Suppose $f : R \rightarrow R$, f and f' are continuous, f'' is piecewise continuous, and $f'' \geq 0$. Then, f is convex.

Proposition 3.2. If $\widetilde{g} \equiv (g_\gamma - g^+) : R \rightarrow R$, then

(a) $\widetilde{g}$ and $\widetilde{g}'$ are continuous and $\widetilde{g}'(t) \geq 0 \quad \forall t \geq 0$.
(b) $\widetilde{g}''$ is piecewise continuous and $\widetilde{g}''(t) \geq 0 \quad \forall t$.

Corollary 3.1. $\tilde{g} \equiv (g_\gamma - g^+) : R \to R$ is convex (by Propositions 3.1 and 3.2).

Proposition 3.3. If $f : R^n \to R$ is convex and $L : R^m \to R^n$ is linear, then $f \circ L$ is convex.

Proposition 3.4. If $f : R \to R$ is convex and non-decreasing for $t \geq 0$ then, $f \circ \|\cdot\|_W : R^n \to R$ is convex.

Corollary 3.2. $\tilde{g} \circ \|\cdot\|_W : R^c \to R$ is convex (by Corollary 3.1 and Propositions 3.2 and 3.4).

Corollary 3.3. $\tilde{g} \circ \|\cdot\|_W \circ L_k : R^{Nc} \to R$ is convex $\forall k = 1,2,\ldots,\ell$ (by Proposition 3.3 and Corollary 3.2).

Corollary 3.4. $F_\gamma - F^+ : R^{Nc} \to R$ is convex (because it is a sum of functions of the form $\tilde{g} \circ \|\cdot\|_W \circ L_k$, which are convex by Corollary 3.3).

Proposition 3.5. The Hessian $\mathbf{D}^2 F^+$ satisfies $\mathbf{D}^2 F^+ = 2 \cdot \overline{W} - \gamma \sum_{k=1}^{\ell} (\mathbf{D}L_k)^T W \, \mathbf{D}L_k$, where the constant matrix $\mathbf{D}L_k$ denotes the total derivative of L_k $(k = 1,2,\ldots,\ell)$, $\overline{W} \equiv \mathit{Block\ diag}(\underbrace{W, W, \cdots, W}_{N-times})$, and $W = diag(w_1, w_2, \cdots, w_c)$.

Proposition 3.6. The Hessian $\mathbf{D}^2 F^+$ is positive semidefinite for γ small enough.

Corollary 3.5. $F^+ : R^{Nc} \to R$ is convex for γ small enough (because for γ small enough $\mathbf{D}^2 F^+$ is positive semidefinite and continuous).

Corollary 3.6. $F_\gamma : R^{Nc} \to R$ is convex for γ small enough (because by Corollaries 3.4 and 3.5, $F_\gamma = (F_\gamma - F^+) + F^+$ is a sum of convex functions for γ small enough).

Proposition 3.7. If $L_k : R^{Nc} \to R^c$ are given by Definition 3.7, then

$$\sum_{k=1}^{N} (\mathbf{D}L_k)^T W\, \mathbf{D}L_k = \overline{W}^{\frac{1}{2}} \Lambda^T \Lambda \overline{W}^{\frac{1}{2}},$$

where Λ is an $Nc \times Nc$ "block circulant" matrix with blocks of size $c \times c$:

$$\Lambda = \begin{bmatrix} -\mathrm{I} & \mathrm{I} & 0 & \cdots & \cdots & 0 \\ 0 & -\mathrm{I} & \mathrm{I} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & -\mathrm{I} & \mathrm{I} \\ \mathrm{I} & 0 & \cdots & \cdots & 0 & -\mathrm{I} \end{bmatrix}.$$

Proposition 3.8. If $\ell = N$, $L_k : R^{Nc} \to R^c$ are given by Definition 3.7, and $\gamma = \frac{1}{2}$, then γ satisfies Proposition 3.6, and Corollaries 3.5 and 3.6.

3.3. THE PROOFS

Proof of Proposition 3.1

Since $f'' \geq 0$ and is piecewise continuous, the following holds for $s < t$

$$0 \leq \int_s^t f''(x)dx = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f''(x)dx = \sum_{i=1}^{n} \left[f'(x_i^-) - f'(x_{i-1}^+) \right],$$

where $s = x_0 < x_1 < \cdots < x_n = t$ and the set $\{x_i : i = 0, 1, \cdots, n\}$ contains the points of discontinuity of f'' inside the interval $[s,t]$. But f' is continuous, so $f'(x_i^-) = f'(x_i^+)$ for all i. Consequently $\int_s^t f''(x)dx = f'(t) - f'(s) \geq 0$ for all $s < t$. This result and the fact that f is continuous lead to the following two conclusions:

$$\text{If } x < y \text{ then } f(y) - f(x) = \int_x^y f'(t)dt \geq \int_x^y f'(x)dt = f'(x)[y - x].$$

$$\text{If } x > y \text{ then } f(x) - f(y) = \int_y^x f'(t)dt \leq \int_y^x f'(x)dt = f'(x)[x - y].$$

The two inequalities above imply $f(y) \geq f(x) + f'(x)[y - x]$ for all $x, y \in R$. Therefore, f is convex by Reference 4, Proposition 4, page 178.

Proof of Proposition 3.2

(a) Since $\tilde{g} = g_\gamma - g^+$ and $g^+ \in C^\infty$, we must show that g_γ and $g_\gamma^{'}$ are continuous. The even function g_γ is continuous on R if it is continuous on $[0, \infty)$, which it clearly is, except possibly at q and r. Continuity of g_γ at q is established by showing that $\lambda^2 t^2 = \alpha - \frac{1}{2}\gamma(t - r)^2$ at $t = q$. Elementary manipulations verify the above identity. Continuity of g_γ at r is established by showing that $\alpha = \alpha - \frac{1}{2}\gamma(t - r)^2$ at $t = r$, which is obvious. Thus, $\tilde{g}$ is continuous.

Next, differentiate g_γ.

$$g_\gamma^{'}(t) = \begin{cases} 2\lambda^2 t & \text{if } |t| < q \\ \gamma(r - t) & \text{if } q \leq t \leq r \\ -\gamma(r + t) & \text{if } -r \leq t \leq -q \\ 0 & \text{if } |t| > r. \end{cases} \tag{3.3.1}$$

Since $g_\gamma^{'}$ is an odd function, it suffices to establish continuity at q and r. Continuity at r is established by showing that $\gamma(r - t) = 0$ at $t = r$, which is obvious. Continuity at q is established by showing that $2\lambda^2 t = \gamma(r - t)$ at $t = q$. Simple algebra and the fact that $r^2 = \alpha(\frac{2}{\gamma} + \frac{1}{\lambda^2})$ and $q = \frac{\alpha}{\lambda^2 r}$ verifies the above identity. Thus, $\tilde{g}'$ is continuous.

That $\tilde{g}'(t) \geq 0$ for $t \geq 0$ is clear from Equation 3.3.1 and the fact that $(g^+)'(t) = -\gamma t$ $(t \in R)$.

(b) Differentiating Equation 3.3.1 yields

$$g_\gamma''(t) = \begin{cases} 2\lambda^2 & \text{if } |t| < q \\ -\gamma & \text{if } q \leq |t| \leq r \\ 0 & \text{if } |t| > r. \end{cases}$$

Since $(g^+)''(t) = -\gamma \quad (t \in R)$, we have

$$\tilde{g}''(t) = \begin{cases} 2\lambda^2 + \gamma & \text{if } |t| < q \\ 0 & \text{if } q \leq |t| \leq r \\ \gamma & \text{if } |t| > r. \end{cases}$$

Thus, $\tilde{g}''$ is piecewise continuous and $\tilde{g}''(t) \geq 0$ for all t.

Proof of Proposition 3.3

Choose $x, y \in R^m$ and $0 \leq t \leq 1$. Then, since L is linear and f is convex we have

$$f \circ L((1-t)x + ty) = f((1-t)Lx + tLy)) \leq$$

$$(1-t)f(Lx) + tf(Ly) = (1-t) \cdot f \circ L(x) + t \cdot f \circ L(y).$$

Thus, the composition $f \circ L$ is convex.

Proof of Proposition 3.4

Set $a = \|(1-t)x + ty\|_W$ and $b = (1-t)\|x\|_W + t\|y\|_W$, $(x, y \in R^n, t \in [0,1])$.
Since $0 \le a \le b$ and f is non-decreasing for $t \ge 0$ and convex,

$$f(a) \le f(b) \le (1-t) \cdot f(\|x\|_W) + t \cdot f(\|y\|_W).$$

Thus, $f \circ \|\cdot\|_W : R^n \to R$ is convex.

Proof of Proposition 3.5

Recall that $F^+ = D + G^+$, $D(u) \equiv \sum_{k=1}^{N} \|u_k - x_k\|_W^2$, $G^+(u) \equiv \sum_{k=1}^{\ell} g^+(\|L_k(u)\|_W)$, $u \equiv [\, u_1{}^T \vdots u_2{}^T \vdots \cdots \vdots u_N{}^T]^T \in R^{Nc}$, and $u_k \equiv [u_{k,1} \;\; u_{k,2} \;\; \cdots \;\; u_{k,c}]^T \in R^c$ for $k = 1, 2, 3, \cdots, N$. So,

$$\frac{\partial D(u)}{\partial u_{ij}} = \frac{\partial}{\partial u_{ij}} \sum_{k=1}^{N} \sum_{m=1}^{c} w_m (u_{km} - x_{km})^2 = 2w_j (u_{ij} - x_{ij}) \quad \Rightarrow$$

$$\frac{\partial^2 D(u)}{\partial u_{rs} \partial u_{ij}} = \begin{cases} 0 & \text{if } rs \ne ij \\ 2w_j & \text{if } rs = ij \end{cases}, \quad (1 \le r, i \le N, 1 \le s, j \le c).$$

Thus, the Hessian $\mathbf{D}^2 D$ of D satisfies $\mathbf{D}^2 D = 2 \cdot \overline{W}$, where $\overline{W} \equiv \text{Block } diag(\underbrace{W, W, \cdots, W}_{N-times})$ is an $Nc \times Nc$ - matrix with $W = diag(w_1, w_2, \cdots, w_c)$.

Next, we compute the Hessian $\mathbf{D}^2 G^+$ of G^+. For this we need the partial derivatives of the compositions $\tilde{g}^+ \circ L_k$, where $\tilde{g}^+$ denotes the composition $g^+ \circ \|\cdot\|_W$; that is,

$\tilde{g}^+(x) = g^+(\|x\|_W) = -\frac{1}{2}\gamma\|x\|_W^2 = -\frac{1}{2}\gamma\sum_{i=1}^{c} w_i x_i^2 \quad (x \in R^c)$ and $G^+(u) \equiv \sum_{k=1}^{\ell} \tilde{g}^+(L_k(u))$ $(u \in R^{Nc})$.

Let $\nabla\tilde{g}^+(t)$ denote the gradient of $\tilde{g}^+ : R^c \to R$ evaluated at $t \in R^c$ and let $\mathbf{D}L_k$ denote the total derivative of $L_k : R^{Nc} \to R^c$ $(1 \le k \le \ell)$. If $t = [t_1\ t_2 \cdots t_c]^T \in R^c$, then

$$\nabla\tilde{g}^+(t) = \left[\frac{\partial\tilde{g}^+}{\partial t_1}(t), \frac{\partial\tilde{g}^+}{\partial t_2}(t), \cdots, \frac{\partial\tilde{g}^+}{\partial t_c}(t)\right] = -\gamma[w_1t_1, w_2t_2, \cdots, w_ct_c].$$

If $L_k(u) = \begin{bmatrix} L_{k1}(u) \\ L_{k2}(u) \\ \vdots \\ L_{kc}(u) \end{bmatrix}$, then $\mathbf{D}L_k = \begin{bmatrix} \frac{\partial L_{k1}}{\partial u_{11}} & \frac{\partial L_{k1}}{\partial u_{12}} & \cdots & \frac{\partial L_{k1}}{\partial u_{Nc}} \\ \frac{\partial L_{k2}}{\partial u_{11}} & \frac{\partial L_{k2}}{\partial u_{12}} & \cdots & \frac{\partial L_{k2}}{\partial u_{Nc}} \\ \vdots & & \ddots & \\ \frac{\partial L_{kc}}{\partial u_{11}} & \frac{\partial L_{kc}}{\partial u_{12}} & \cdots & \frac{\partial L_{kc}}{\partial u_{Nc}} \end{bmatrix}$ $(1 \le k \le \ell)$.

Note that $\nabla\tilde{g}^+(t)$ is a $1 \times c$ - matrix and $\mathbf{D}L_k$ is a $c \times Nc$ - matrix. Since L_k is linear, $\mathbf{D}L_k$ is a constant matrix $(1 \le k \le \ell)$.

The gradient of the composition $\tilde{g}^+ \circ L_k$ is obtained using the chain rule:

$$\nabla(\tilde{g}^+ \circ L_k)(u) = \left(\nabla\tilde{g}^+\Big|_{t = L_k(u)}\right) \cdot \mathbf{D}L_k, \quad (u \in R^{Nc}). \tag{3.3.2}$$

Since, by definition, $\nabla(\tilde{g}^+ \circ L_k)(u) \equiv \left[\frac{\partial}{\partial u_{11}}\tilde{g}^+(L_k(u)), \frac{\partial}{\partial u_{12}}\tilde{g}^+(L_k(u)), \cdots, \frac{\partial}{\partial u_{Nc}}\tilde{g}^+(L_k(u))\right]$, it follows from Equation 3.3.2 that

$$\frac{\partial}{\partial u_{ij}}\tilde{g}^{+}(L_k(u))=\sum_{m=1}^{c}\frac{\partial \tilde{g}^{+}}{\partial t_m}(L_k(u))\cdot\frac{\partial L_{km}}{\partial u_{ij}},\quad (u\in R^{Nc},\quad (i,j)\in\{(1,1),(1,2),\cdots(N,c)\},$$

$1\leq k\leq \ell$).

Similarly, since $\frac{\partial \tilde{g}^{+}}{\partial t_m}:R^c\to R$,

$$\frac{\partial}{\partial u_{rs}}\left(\frac{\partial \tilde{g}^{+}}{\partial t_m}(L_k(u))\right)=\sum_{n=1}^{c}\frac{\partial^2 \tilde{g}^{+}}{\partial t_n\partial t_m}(L_k(u))\cdot\frac{\partial L_{kn}}{\partial u_{rs}}, \tag{3.3.3}$$

where ($u\in R^{Nc}$, $(r,s)\in\{(1,1),(1,2),\cdots(N,c)\}$, $1\leq k\leq \ell$).

Finally, $\frac{\partial}{\partial u_{ij}}G^{+}(u)\equiv\sum_{k=1}^{\ell}\frac{\partial}{\partial u_{ij}}\tilde{g}^{+}(L_k(u))=\sum_{k=1}^{\ell}\sum_{m=1}^{c}\frac{\partial \tilde{g}^{+}}{\partial t_m}(L_k(u))\cdot\frac{\partial L_{km}}{\partial u_{ij}}$ and, since $\frac{\partial L_{km}}{\partial u_{ij}}$ is a constant, $\frac{\partial^2}{\partial u_{rs}\partial u_{ij}}G^{+}(u)=\sum_{k=1}^{\ell}\sum_{m=1}^{c}\frac{\partial}{\partial u_{rs}}\left[\frac{\partial \tilde{g}^{+}}{\partial t_m}(L_k(u))\right]\cdot\frac{\partial L_{km}}{\partial u_{ij}}$.

Applying Equation 3.3.3 leads to

$$\frac{\partial^2}{\partial u_{rs}\partial u_{ij}}G^{+}(u)=\sum_{k=1}^{\ell}\sum_{m=1}^{c}\sum_{n=1}^{c}\frac{\partial^2 \tilde{g}^{+}}{\partial t_n\partial t_m}(L_k(u))\cdot\frac{\partial L_{kn}}{\partial u_{rs}}\cdot\frac{\partial L_{km}}{\partial u_{ij}},\quad (u\in R^{Nc},\ (r,s)\ \text{and}$$

$(i,j)\in\{(1,1),(1,2),\cdots(N,c)\}$).

Next, since

$$\frac{\partial \tilde{g}^+}{\partial t_m}(t) = -\gamma \cdot w_m t_m, \quad (t = [t_1 \; t_2 \cdots t_c]^T \in R^c, \; 1 \le m \le c),$$

$$\frac{\partial^2 \tilde{g}^+}{\partial t_n \partial t_m}(t) = -\gamma \cdot w_m \cdot \delta_{nm} \equiv \begin{cases} -\gamma \cdot w_m & \text{if } n = m \\ 0 & \text{if } n \ne m. \end{cases}$$

Hence,

$$\frac{\partial^2}{\partial u_{rs} \partial u_{ij}} G^+(u) = \sum_{k=1}^{\ell} \sum_{m=1}^{c} -\gamma \cdot w_m \cdot \frac{\partial L_{km}}{\partial u_{rs}} \cdot \frac{\partial L_{km}}{\partial u_{ij}},$$

where ($u \in R^{Nc}$, (r,s) and (i,j) $\in \{(1,1), (1,2), \cdots (N,c)\}$), which implies that

$$\mathbf{D}^2 G^+ = -\gamma \cdot \sum_{k=1}^{\ell} (\mathbf{D}L_k)^T W \, \mathbf{D}L_k .$$

Adding $\mathbf{D}^2 D$ and $\mathbf{D}^2 G^+$ completes the proof.

Proof of Proposition 3.6

It will be shown that, for γ small enough, $u^T \mathbf{D}^2 F^+ u \ge 0 \;\; \forall u \in R^{Nc}$. The norm of the linear mappings $\mathbf{D}L_k : R^{Nc} \to R^c$ will be instrumental. These norms are defined as

$$\|\mathbf{D}L_k\| \equiv \sup_{\|u\|_{\overline{W}}=1} \|\mathbf{D}L_k u\|_W ,$$

where $\|u\|_{\overline{W}}^2 \equiv u^T \overline{W} u = \sum_{k=1}^{N} \|u_k\|_W^2$ and $\|u_k\|_W^2 \equiv u_k^T W u_k$, $(k = 1, 2, \cdots, N)$ (see Definition 3.6).

$$u^T \mathbf{D}^2 F^+ u = u^T \left[2 \cdot \overline{W} - \gamma \sum_{k=1}^{\ell} (\mathbf{D}L_k)^T W \, \mathbf{D}L_k \right] u = 2u^T \overline{W} u - \gamma \sum_{k=1}^{\ell} u^T (\mathbf{D}L_k)^T W \, \mathbf{D}L_k u$$

$$= 2 \cdot \|u\|_{\overline{W}}^2 - \gamma \sum_{k=1}^{\ell} \|\mathbf{D}L_k u\|_W^2 \geq 2 \cdot \|u\|_{\overline{W}}^2 - \gamma \sum_{k=1}^{\ell} \|\mathbf{D}L_k\|^2 \|u\|_{\overline{W}}^2$$

$$= \left[2 - \gamma \sum_{k=1}^{\ell} \|\mathbf{D}L_k\|^2 \right] \cdot \|u\|_{\overline{W}}^2 .$$

Therefore, if $\gamma \leq \dfrac{2}{\sum_{k=1}^{\ell} \|\mathbf{D}L_k\|^2}$, then $\left[2 - \gamma \sum_{k=1}^{\ell} \|\mathbf{D}L_k\|^2 \right] \geq 0$ and $u^T \mathbf{D}^2 F^+ u \geq 0 \;\; \forall u \in R^{Nc}$.

Hence, $\mathbf{D}^2 F^+$ is positive semidefinite for γ small enough.

Proof of Proposition 3.7

First we'll show that $\mathbf{D}L_N$ is a block matrix of the form

$$\mathbf{D}L_N = [\, \mathrm{I} \,\vdots\, 0 \,\vdots \cdots \vdots\, 0 \,\vdots\, -\mathrm{I} \,] \tag{3.3.4}$$

and

$$\mathbf{D}L_k = [\, 0 \,\vdots \cdots \vdots\, 0 \,\vdots\, -\mathrm{I} \,\vdots\, \mathrm{I} \,\vdots\, 0 \,\vdots \cdots \vdots\, 0 \,] \quad \text{for } k = 1, 2, \cdots, N-1, \tag{3.3.5}$$

where $-\mathrm{I}$ appears in the k-th block, I in the $(k+1)$-st block, I denotes the $c \times c$ identity matrix, and there are N $c \times c$-blocks.

The n-th component of $L_N : R^{Nc} \rightarrow R^c$ is given by

$$L_{Nn}(u) = u_{1,n} - u_{N,n} \qquad (n = 1, 2, \cdots, c).$$

Thus, the $[n,(i,j)]$-th entry of $\mathbf{D}L_N$, denoted by $(\mathbf{D}L_N)_{n,(i,j)}$ is given by

$$(\mathbf{D}L_N)_{n,(i,j)} \equiv \frac{\partial L_{Nn}}{\partial u_{ij}} = \begin{cases} 0 & \text{if } i \neq 1, N \\ 0 & \text{if } j \neq n \\ -1 & \text{if } i = N \text{ and } n = j \\ 1 & \text{if } i = 1 \text{ and } n = j \end{cases} \qquad (1 \leq n \leq c, \quad 1 \leq i \leq N, \quad 1 \leq j \leq c). \tag{3.3.6}$$

The i-th $c \times c$-block of $\mathbf{D}L_N$ is denoted by $(\mathbf{D}L_N)_{*,(i,*)}$ $(1 \leq i \leq N)$. Its (n,j)-entry is $\frac{\partial L_{Nn}}{\partial u_{ij}}$ $(1 \leq n \leq c, 1 \leq j \leq c)$. By Equation 3.3.6, $(\mathbf{D}L_N)_{*,(i,*)}$ satisfies

$$(\mathbf{D}L_N)_{*,(i,*)} = \begin{cases} 0 & \text{if } i \neq 1, N \\ \mathrm{I} & \text{if } i = 1 \\ -\mathrm{I} & \text{if } i = N. \end{cases}$$

This establishes Equation 3.3.4.

For $1 \leq k \leq N-1$ the n-th component of $L_k : R^{Nc} \rightarrow R^c$ is given by

$$L_{kn}(u) = u_{k+1,n} - u_{k,n} \qquad (n = 1, 2, \cdots, c).$$

Thus, the $[n,(i,j)]$-th entry of $\mathbf{D}L_k$, denoted by $(\mathbf{D}L_k)_{n,(i,j)}$ is given by

$$(\mathbf{D}L_k)_{n,(i,j)} \equiv \frac{\partial L_{kn}}{\partial u_{ij}} = \begin{cases} 0 & \text{if } i \neq k, k+1 \\ 0 & \text{if } j \neq n \\ -1 & \text{if } i = k \text{ and } n = j \\ 1 & \text{if } i = k+1 \text{ and } n = j \end{cases} \qquad (1 \leq n \leq c, \quad 1 \leq i \leq N, \quad 1 \leq j \leq c).$$

(3.3.7)

The i-th $c \times c$-block of $\mathbf{D}L_k$ is denoted by $(\mathbf{D}L_k)_{*,(i,*)}$ $(1 \leq i \leq N)$. Its (n,j)-entry is $\frac{\partial L_{kn}}{\partial u_{ij}}$ $(1 \leq n \leq c, 1 \leq j \leq c)$. By Equation 3.3.7, $(\mathbf{D}L_k)_{*,(i,*)}$ satisfies

$$(\mathbf{D}L_k)_{*,(i,*)} = \begin{cases} 0 & \text{if } i \neq k, k+1 \\ \mathrm{I} & \text{if } i = k+1 \\ -\mathrm{I} & \text{if } i = k. \end{cases}$$

This establishes Equation 3.3.5.

Now, the product $(\mathbf{D}L_N)^T W \, \mathbf{D}L_N$ clearly has the form

$$(\mathbf{D}L_N)^T W \, \mathbf{D}L_N = \begin{bmatrix} \mathrm{I} \\ 0 \\ \vdots \\ 0 \\ -\mathrm{I} \end{bmatrix} W \begin{bmatrix} \mathrm{I} & 0 & \cdots & 0 & -\mathrm{I} \end{bmatrix} = \begin{bmatrix} W & 0 & \cdots & 0 & -W \\ 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \\ -W & 0 & \cdots & 0 & W \end{bmatrix}$$

and

$$(\mathbf{D}L_k)^T W\, \mathbf{D}L_k = \left[\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & W & -W & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -W & W & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \end{array}\right] \qquad (1 \le k \le N-1),$$

where W appears in the (k,k)- and $(k+1,k+1)$-block, and $-W$ appears in the $(k+1,k)$- and $(k,k+1)$-block. Consequently,

$$\sum_{k=1}^{N} (\mathbf{D}L_k)^T W\, \mathbf{D}L_k = \begin{bmatrix} 2W & -W & 0 & \cdots & 0 & -W \\ -W & 2W & -W & 0 & \cdots & 0 \\ 0 & -W & 2W & -W & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -W & 2W & -W \\ -W & 0 & \cdots & 0 & -W & 2W \end{bmatrix}.$$

Since $\overline{W}^{\frac{1}{2}} \Lambda^T \Lambda \overline{W}^{\frac{1}{2}}$ has the same form, it follows that $\sum_{k=1}^{N} (\mathbf{D}L_k)^T W\, \mathbf{D}L_k = \overline{W}^{\frac{1}{2}} \Lambda^T \Lambda \overline{W}^{\frac{1}{2}}$.

Proof of Proposition 3.8

By Propositions 3.5 and 3.7 we can write $\mathbf{D}^2F^+$ as $\mathbf{D}^2F^+ = 2\cdot\overline{W} - \gamma\overline{W}^{\frac{1}{2}}\Lambda^T\Lambda\overline{W}^{\frac{1}{2}}$. So, as in the proof of Proposition 3.6,

$$\begin{aligned} u^T\mathbf{D}^2F^+u &= u^T\left[2\cdot\overline{W} - \gamma\overline{W}^{\frac{1}{2}}\Lambda^T\Lambda\overline{W}^{\frac{1}{2}}\right]u = 2u^T\overline{W}u - \gamma\cdot u^T\overline{W}^{\frac{1}{2}}\Lambda^T\Lambda\overline{W}^{\frac{1}{2}}u \\ &= 2\cdot\left\|\overline{W}^{\frac{1}{2}}u\right\|^2 - \gamma\left\|\Lambda\overline{W}^{\frac{1}{2}}u\right\|^2 \geq 2\cdot\left\|\overline{W}^{\frac{1}{2}}u\right\|^2 - \gamma\|\Lambda\|^2\left\|\overline{W}^{\frac{1}{2}}u\right\|^2 = \\ &= \left[2-\gamma\|\Lambda\|^2\right]\cdot\left\|\overline{W}^{\frac{1}{2}}u\right\|^2. \end{aligned}$$

Since $\gamma = \frac{1}{2}$, $u^T\mathbf{D}^2F^+u \geq 0 \quad \forall u \in R^{Nc}$ provided $\|\Lambda\|^2 \leq 4$. But $\|\Lambda\|^2$ equals the largest eigenvalue of $\Lambda^T\Lambda$. Thus, the proof will be complete once we show that the eigenvalues of $\Lambda^T\Lambda$ are bounded by 4.

The matrix $\Lambda^T\Lambda$ is a "block circulant" matrix:

$$\Lambda^T\Lambda = \begin{bmatrix} 2\mathrm{I} & -\mathrm{I} & 0 & \cdots & 0 & -\mathrm{I} \\ -\mathrm{I} & 2\mathrm{I} & -\mathrm{I} & 0 & \cdots & 0 \\ 0 & -\mathrm{I} & 2\mathrm{I} & -\mathrm{I} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\mathrm{I} & 2\mathrm{I} & -\mathrm{I} \\ -\mathrm{I} & 0 & \cdots & 0 & -\mathrm{I} & 2\mathrm{I} \end{bmatrix}.$$

Since the blocks of $\Lambda^T\Lambda$ are multiples of the identity matrix, $\Lambda^T\Lambda$ has the same eigenvalues as the circulant matrix

$$A \equiv \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & -1 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 2 & -1 \\ -1 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix},$$

which are given by the following expression.

Fact 3.8.1. If A is a circulant matrix, then the eigenvalues of A are given by (Reference 5)

$$\lambda_m = \sum_{r=1}^{N} A_{1r} e^{2\pi i \frac{m}{N}(r-1)} \qquad m = 1, 2, \cdots, N.$$

Thus, the eigenvalues of the matrix A above are given by

$$\lambda_m = 2 - e^{2\pi i \frac{m}{N}} - e^{-2\pi i \frac{m}{N}} = 2[1 - \cos(2\pi \tfrac{m}{N})], \quad m = 1, 2, \cdots, N. \tag{3.3.8}$$

Since $-1 \le \cos(2\pi \frac{m}{N}) \le 1$ for all m, $\lambda_m \le 4$ for all m. Note that if N is even, then $\lambda_m = 4$ when $m = \frac{N}{2}$. If N is odd and $m = \frac{N+1}{2}$, then $\lambda_m = 2[1 + \cos(\frac{\pi}{N})] \to 4$ as $N \to \infty$. Thus, the upper bound 4 in general cannot be reduced.

4. THE 2D-MULTICHANNEL GNC CASE

4.1. DEFINITIONS

4.1. $D: R^{nmc} \to R$ is defined as $D(u) \equiv \sum_{\substack{1\le i\le n\\1\le j\le m}} \left\| u_{ij} - x_{ij} \right\|_W^2$, where

$$u_{ij} \equiv [u_{ij,1} \;\; u_{ij,2} \;\; \cdots \;\; u_{ij,c}]^T \in R^c \quad (1 \le i \le n, 1 \le j \le m),$$

$$u \equiv [\, u_{11}{}^T \vdots\, u_{12}{}^T \vdots \cdots \vdots\, u_{1m}{}^T \vdots\, u_{21}{}^T \vdots\, u_{22}{}^T \vdots \cdots \vdots\, u_{2m}{}^T \vdots \cdots \vdots\, u_{n1}{}^T \vdots\, u_{n2}{}^T \vdots \cdots \vdots\, u_{nm}{}^T \,]^T \in R^{nmc},$$

$$W = diag(w_1, w_2, \cdots, w_c) \quad (w_i > 0),$$

and the weighted norm $\| \; \|_W$ is defined as $\|y\|_W^2 \equiv \sum_{i=1}^{c} w_i y_i^2$.

4.2. $M_{ij} : R^{nmc} \to R^c$ is defined as $M_{ij}(u) \equiv u_{i,j+1} - u_{i,j}$ for $1 \le i \le n, 1 \le j \le m-1$ and $M_{im}(u) \equiv u_{i,1} - u_{i,m}$ for $1 \le i \le n$ and $u \in R^{nmc}$ as defined in **4.1**.

$N_{ij} : R^{nmc} \to R^c$ is defined as $N_{ij}(u) \equiv u_{i+1,j} - u_{i,j}$ for $1 \le i \le n-1, 1 \le j \le m$ and $N_{nj}(u) \equiv u_{1,j} - u_{n,j}$ for $1 \le j \le m$ and $u \in R^{nmc}$ as defined in **4.1**.

4.3. $\widetilde{G} : R^{nmc} \times \{0,1\}^{nm} \times \{0,1\}^{nm} \to R$ is given by $\widetilde{G}(u,a,b) = \widetilde{H}(u,a) + \widetilde{V}(u,b)$, where

$$\widetilde{H}(u,a) \equiv \sum_{\substack{1\le i\le n\\1\le j\le m}} h_{\lambda_1,\alpha_1}(\|M_{ij}(u)\|_W, a_{ij}), \quad \widetilde{V}(u,b) \equiv \sum_{\substack{1\le i\le n\\1\le j\le m}} h_{\lambda_2,\alpha_2}(\|N_{ij}(u)\|_W, b_{ij}),$$

$$a \equiv [a_{11} \; a_{12} \cdots a_{1m} \; a_{21} \; a_{22} \cdots a_{2m} \cdots a_{n1} \; a_{n2} \cdots a_{nm}]^T \in \{0,1\}^{nm},$$

$$b \equiv [b_{11}\ b_{12} \cdots b_{1m}\ b_{21}\ b_{22} \cdots b_{2m} \cdots b_{n1}\ b_{n2} \cdots b_{nm}]^T \in \{0,1\}^{nm},\ u \in R^{nmc}$$

as defined in **4.1.**, and $h_{\lambda,\alpha} : R \times \{0,1\} \to R$ is defined by

$$h_{\lambda,\alpha}(t,b) \equiv \lambda^2 t^2 (1-b) + \alpha b \qquad (t \in R, b \in \{0,1\}).$$

4.4. $\widetilde{F} : R^{nmc} \times \{0,1\}^{nm} \times \{0,1\}^{nm} \to R$ is given by $\widetilde{F} \equiv D + \widetilde{G}$.

4.5. $G : R^{nmc} \to R$ is given by $G(u) = H(u) + V(u)$, where

$$H(u) \equiv \sum_{\substack{1 \le i \le n \\ 1 \le j \le m}} g_{\lambda_1,\alpha_1}(\|M_{ij}(u)\|_W), \qquad V(u) \equiv \sum_{\substack{1 \le i \le n \\ 1 \le j \le m}} g_{\lambda_2,\alpha_2}(\|N_{ij}(u)\|_W), \quad u \in R^{nmc}$$

as defined in **4.1.**, and $g_{\lambda,\alpha} : R \to R$ is defined by

$$g_{\lambda,\alpha}(t) \equiv \begin{cases} \lambda^2 t^2 & \text{if } |t| \le \sqrt{\alpha}/\lambda \\ \alpha & \text{if } |t| \ge \sqrt{\alpha}/\lambda \end{cases}, \qquad \text{for } \alpha > 0.$$

4.6. $G_\gamma : R^{nmc} \to R$ is defined as G with g_{λ_1,α_1} and g_{λ_2,α_2} replaced by $g_{\gamma,\lambda_1,\alpha_1}$ and $g_{\gamma,\lambda_2,\alpha_2}$, respectively, and $g_{\gamma,\lambda,\alpha} : R \to R$ is defined for $\alpha > 0$ by

$$g_{\gamma,\lambda,\alpha}(t) \equiv \begin{cases} \lambda^2 t^2 & \text{if } |t| \le q \\ \alpha - \frac{1}{2}\gamma(|t| - r)^2 & \text{if } q \le |t| \le r \\ \alpha & \text{if } \mathrm{r} \le |t| \end{cases},$$

$\gamma > 0$ and $q = \dfrac{\alpha}{\lambda^2 r}$, $r^2 = \alpha(\dfrac{2}{\gamma} + \dfrac{1}{\lambda^2})$.

4.7. $G^+ : R^{nmc} \to R$ is defined as G with g_{λ_1,α_1} and g_{λ_2,α_2} both replaced by $g^+ : R \to R$ defined by $g^+(t) \equiv -\frac{1}{2}\gamma t^2$ $(t \in R)$.

4.8. F, F_γ, and $F^+ : R^{nmc} \to R$ are defined, respectively, as $F = D + G$, $F_\gamma = D + G_\gamma$, and $F^+ = D + G^+$.

4.2. CONVEXITY OF $F_\gamma : R^{nmc} \to R$

Proposition 4.1. $F_\gamma - F^+ : R^{nmc} \to R$ is convex (proof: same as Corollary 3.4).

Proposition 4.2. The Hessian $\mathbf{D}^2 F^+$ satisfies

$$\mathbf{D}^2 F^+ = 2 \cdot \overline{W}_{nm} - \gamma \cdot \left[\sum_{\substack{1 \le i \le n \\ 1 \le j \le m}} (\mathbf{D}M_{ij})^T W \, \mathbf{D}M_{ij} + \sum_{\substack{1 \le i \le n \\ 1 \le j \le m}} (\mathbf{D}N_{ij})^T W \, \mathbf{D}N_{ij} \right], \tag{4.2.1}$$

where the constant matrices $\mathbf{D}M_{ij}$ and $\mathbf{D}N_{ij}$ denote, respectively, the total derivatives of M_{ij} and N_{ij} $(1 \le i \le n, 1 \le j \le m)$, $\overline{W}_{nm} \equiv Block\ Diag(\underbrace{W, W, \cdots, W}_{nm-times})$, and $W = diag(w_1, w_2, \cdots, w_c)$.

Proposition 4.3. If $M_{ij} : R^{nmc} \to R^c$ and $N_{ij} : R^{nmc} \to R^c$ are given by Definition 4.2, then

$$\sum_{\substack{1\le i\le n\\1\le j\le m}} (\mathbf{D}M_{ij})^T W\,\mathbf{D}M_{ij} = \overline{W}_{nm}^{\frac{1}{2}} \cdot \Omega^T \Omega \cdot \overline{W}_{nm}^{\frac{1}{2}}\,, \tag{4.2.a}$$

where $\Omega \equiv Block\ Diag(\underbrace{\Lambda_m, \Lambda_m, \cdots, \Lambda_m}_{n-times})$ and Λ_m is the $mc \times mc$ "block circulant" matrix

$$\Lambda_m = \begin{bmatrix} -I & I & 0 & \cdots & \cdots & 0 \\ 0 & -I & I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & -I & I \\ I & 0 & \cdots & \cdots & 0 & -I \end{bmatrix}, \quad \text{with blocks of size } c \times c.$$

$$\sum_{\substack{1\le i\le n\\1\le j\le m}} (\mathbf{D}N_{ij})^T W\,\mathbf{D}N_{ij} = \overline{W}_{nm}^{\frac{1}{2}} \cdot \Lambda_n^T \Lambda_n \cdot \overline{W}_{nm}^{\frac{1}{2}}, \tag{4.2.b}$$

where Λ_n is a "block circulant" matrix of the same form as Λ_m of size $nmc \times nmc$ and blocks of size $mc \times mc$.

Proposition 4.4. If $M_{ij} : R^{nmc} \to R^c$ and $N_{ij} : R^{nmc} \to R^c$ are given by Definition 4.2 and $\gamma = \frac{1}{4}$, then the Hessian $\mathbf{D}^2 F^+$ is positive semidefinite. Thus, F^+ is convex.

Corollary 4.1. If $M_{ij} : R^{nmc} \to R^c$ and $N_{ij} : R^{nmc} \to R^c$ are given by Definition 4.2 and $\gamma = \frac{1}{4}$, then F_γ is convex (proof: same as Corollary 3.6).

4.3. THE PROOFS

Proof of Proposition 4.2

Let Ind denote the set of indices $\{(i,j,k) : 1 \le i \le n, 1 \le j \le m, 1 \le k \le c\}$.

$$\frac{\partial D(u)}{\partial u_{rst}} = \frac{\partial}{\partial u_{rst}}\left[\sum_{\substack{1\le i\le n\\1\le j\le m}} \sum_{k=1}^{c} w_k (u_{ijk} - x_{ijk})^2\right] = 2w_t(u_{rst} - x_{rst}), \quad (r,s,t) \in Ind$$

$$\Rightarrow \qquad \frac{\partial^2 D(u)}{\partial u_{ijk}\partial u_{rst}} = \begin{cases} 0 & \text{if} \quad ijk \ne rst \\ 2w_t & \text{if} \quad ijk = rst, \end{cases} \quad (i,j,k), \quad (r,s,t) \in Ind\,.$$

Thus, $\mathbf{D}^2 D$ is an $nmc \times nmc$ block matrix with $(nm)^2$ blocks of size $c \times c$. All the blocks are zero except for the nm diagonal blocks, which are all equal to $2 \cdot W$. Thus, $\mathbf{D}^2 D = 2 \cdot \overline{W}_{nm}$.

Next, we compute $\mathbf{D}^2 H^+$ and $\mathbf{D}^2 V^+$. But H^+ has the same form as G^+ of the 1D-case (see Proposition 3.5), except that it involves M_{ij} and a double sum (sum over $i,j, \ 1 \le i \le n, 1 \le j \le m$) instead of L_k and a sum over $k, \ 1 \le k \le N$. Thus, $\mathbf{D}^2 H^+$ has the same form as $\mathbf{D}^2 G^+$ of the 1D-case, namely

$$\mathbf{D}^2 H^+ = -\gamma \sum_{\substack{1\le i\le n\\1\le j\le m}} (\mathbf{D}M_{ij})^T W\, \mathbf{D}M_{ij}\,.$$

Similarly, $\mathbf{D}^2 V^+ = -\gamma \sum_{\substack{1\le i\le n\\1\le j\le m}} (\mathbf{D}N_{ij})^T W\, \mathbf{D}N_{ij}$. This establishes Equation 4.2.1.

Proof of Proposition 4.3

The total derivatives $\mathbf{D}M_{ij}$ are matrices of dimension $c \times nmc$. The columns of these matrices are indexed by the indices of the components of the variable $u \in R^{nmc}$, namely, the triplets $(r,s,t) \in Ind$ ordered by the rule

$$(i,j,k) < (r,s,t) \quad \Leftrightarrow \quad \begin{array}{llll} i<r, & \text{or} & & \\ i=r & \text{and} & j<s, & \text{or} \\ i=r, & j=s, & \text{and} & k<t. \end{array}$$

Therefore, these matrices (as well as the matrices $\mathbf{D}N_{ij}$) naturally decompose into nm blocks of size $c \times c$. These blocks can naturally be indexed by the tuples $\{ (i,j) : 1 \le i \le n, 1 \le j \le m \}$ ordered by the rule

$$(i,j) < (r,s) \quad \Leftrightarrow \quad \begin{array}{lll} i<r, & \text{or} & \\ i=r & \text{and} & j<s. \end{array}$$

Moreover, since W is a $c \times c$ matrix, the products $(\mathbf{D}M_{ij})^T W \, \mathbf{D}M_{ij}$ are $nmc \times nmc$ matrices that naturally decompose into:

(a) $(nm)^2$ blocks of size $c \times c$, or
(b) n^2 blocks of size $mc \times mc$, or
(c) m^2 blocks of size $nc \times nc$.

This observation also applies to the products $(\mathbf{D}N_{ij})^T W \, \mathbf{D}N_{ij}$. In case (a), the $c \times c$ blocks will be indexed by pairs of tuples $((i,j),(r,s))$, $1 \le i,r \le n$, $1 \le j,s \le m$. In case (b), the $mc \times mc$ blocks will simply be indexed by tuples (i,r), $1 \le i,r \le n$. In case (c), the $nc \times nc$ blocks will be indexed by tuples (j,s), $1 \le j,s \le m$.

First we'll show that $\mathbf{D}M_{im}$ and $\mathbf{D}M_{ij}$ are block matrices of the following form:

$$\mathbf{D}M_{im} = \begin{bmatrix} 0 & \cdots & 0 & \mathrm{I} & 0 & \cdots & 0 & \text{-I} & 0 & \cdots & 0 \end{bmatrix}, \quad (1 \le i \le n) \tag{4.3.1}$$

where I denotes the $c \times c$ identity matrix and appears in the $(i,1)$ - block, -I appears in the (i,m) - block, and there are nm blocks of size $c \times c$. And

$$\mathbf{D}M_{ij} = \begin{bmatrix} 0 & \cdots & 0 & \text{-I} & \mathrm{I} & 0 & \cdots & 0 \end{bmatrix}, \quad (1 \le i \le n,\ 1 \le j \le m-1) \tag{4.3.2}$$

where -I appears in the (i,j) - block, I appears in the $(i,j+1)$ - block, and there are nm blocks of size $c \times c$.

Choose $i \in \{1, 2, \cdots, n\}$. The *k-th* component of $M_{im} : R^{nmc} \to R^c$ is

$$M_{im,k}(u) = u_{i,1,k} - u_{i,m,k} \qquad (1 \le k \le c).$$

Thus the $[k,(r,s,t)]$ entry of $\mathbf{D}M_{im}$, denoted by $(\mathbf{D}M_{im})_{k,(r,s,t)}$, is

$$(\mathbf{D}M_{im})_{k,(r,s,t)} \equiv \frac{\partial M_{im,k}}{\partial u_{r,s,t}} = \begin{cases} 0 & \text{if } r \ne i \\ 0 & \text{if } t \ne k \\ 0 & \text{if } s \ne 1,m \\ 1 & \text{if } r = i, s = 1, t = k \\ -1 & \text{if } r = i, s = m, t = k \end{cases}, \quad (r,s,t) \in Ind,\ 1 \le k \le c. \tag{4.3.3}$$

The (r,s)*-th* $c \times c$ block of $\mathbf{D}M_{im}$ is denoted by $(\mathbf{D}M_{im})_{*,(r,s,*)}$ $(1 \le r \le n,\ \ 1 \le s \le m)$. Its (k,t) entry is $\dfrac{\partial M_{im,k}}{\partial u_{r,s,t}}$ $(1 \le k \le c,\ \ 1 \le t \le c)$. By Equation 4.3.3, $(\mathbf{D}M_{im})_{*,(r,s,*)}$ satisfies

$$(\mathbf{D}M_{im})_{*,(r,s,*)} = \begin{cases} 0 & \text{if } r \ne i \\ 0 & \text{if } s \ne 1, m \\ \mathrm{I} & \text{if } r = i, s = 1 \\ \text{-I} & \text{if } r = i, s = m \end{cases} \quad , \quad (1 \le r \le n,\ \ 1 \le s \le m).$$

This establishes Equation 4.3.1.

Choose $i \in \{1, 2, \cdots, n\}$ and $j \in \{1, 2, \cdots, m-1\}$. The *k-th* component of $M_{ij} : R^{nmc} \to R^c$ is

$$M_{ij,k}(u) = u_{i,j+1,k} - u_{i,j,k} \qquad (1 \le k \le c).$$

Thus the $[k,(r,s,t)]$ entry of $\mathbf{D}M_{ij}$, denoted by $(\mathbf{D}M_{ij})_{k,(r,s,t)}$, is

$$(\mathbf{D}M_{ij})_{k,(r,s,t)} \equiv \frac{\partial M_{ij,k}}{\partial u_{r,s,t}} = \begin{cases} 0 & \text{if } r \ne i \\ 0 & \text{if } t \ne k \\ 0 & \text{if } s \ne j, j+1 \\ 1 & \text{if } r = i, s = j+1, t = k \\ -1 & \text{if } r = i, s = j, t = k \end{cases} \quad , \ (r,s,t) \in Ind,\ 1 \le k \le c. \quad (4.3.4)$$

The (r,s)-*th* $c \times c$ block of $\mathbf{D}M_{ij}$ is denoted by $(\mathbf{D}M_{ij})_{*,(r,s,*)}$ $(1 \le r \le n, \quad 1 \le s \le m)$. Its (k,t) entry is $\dfrac{\partial M_{ij,k}}{\partial u_{r,s,t}}$ $(1 \le k \le c, \quad 1 \le t \le c)$. By (4.3.4), $(\mathbf{D}M_{ij})_{*,(r,s,*)}$ satisfies

$$(\mathbf{D}M_{ij})_{*,(r,s,*)} = \begin{cases} 0 & \text{if } r \ne i \\ 0 & \text{if } s \ne j, j+1 \\ \mathrm{I} & \text{if } r = i, s = j+1 \\ \text{-I} & \text{if } r = i, s = j \end{cases} \qquad (1 \le r \le n, \quad 1 \le s \le m).$$

This establishes Equation 4.3.2.

By Equation 4.3.1 the product $(\mathbf{D}M_{im})^T W \, \mathbf{D}M_{im}$ has the form

$$(\mathbf{D}M_{im})^T W \, \mathbf{D}M_{im} = \left[\begin{array}{ccc|cccccc|ccc} 0 & \cdots & 0 & 0 & & \cdots & & & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & & \cdots & & & 0 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & W & 0 & \cdots & & 0 & -W & 0 & \cdots & 0 \\ & & & 0 & \cdots & \cdots & \cdots & & 0 & & & \\ \vdots & & \vdots & \vdots & & & & & \vdots & \vdots & & \vdots \\ & & & 0 & \cdots & \cdots & \cdots & & 0 & & & \\ 0 & \cdots & 0 & -W & 0 & \cdots & & 0 & W & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & & \cdots & & & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & & \cdots & & & 0 & 0 & \cdots & 0 \end{array}\right] \qquad (1 \le i \le n),$$

where W appears in the $((i,1),(i,1))$ block and in the $((i,m),(i,m))$ block, $-W$ appears in the $((i,1),(i,m))$ block and in the $((i,m),(i,1))$ block, and all other blocks are zero.

By Equation 4.3.2 the product $(\mathbf{D}M_{ij})^T W \, \mathbf{D}M_{ij}$ has the form

$$(\mathbf{D}M_{ij})^T W\, \mathbf{D}M_{ij} = \left[\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & W & -W & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -W & W & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \end{array}\right] \quad (1 \le i \le n,\ 1 \le j \le m-1),$$

where W appears in the $((i,j),(i,j))$ block and the $((i,j+1),(i,j+1))$ block, $-W$ appears in the $((i,j+1),(i,j))$ block and the $((i,j),(i,j+1))$ block, and all other blocks are zero.

Consequently,
$$\sum_{j=1}^{m} (\mathbf{D}M_{ij})^T W\, \mathbf{D}M_{ij} = \left[\begin{array}{c|c|c} 0 & 0 & 0 \\ \hline 0 & \overline{W}_m^{\frac{1}{2}} \Lambda_m^T \Lambda_m \overline{W}_m^{\frac{1}{2}} & 0 \\ \hline 0 & 0 & 0 \end{array}\right],$$

where
$$\overline{W}_m \equiv Block\ Diag(\underbrace{W, W, \cdots, W}_{m-times}),$$

$$\overline{W}_m^{\frac{1}{2}} \Lambda_m^T \Lambda_m \overline{W}_m^{\frac{1}{2}} = \begin{bmatrix} 2W & -W & 0 & \cdots & 0 & -W \\ -W & 2W & -W & 0 & \cdots & 0 \\ 0 & -W & 2W & -W & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -W & 2W & -W \\ -W & 0 & \cdots & 0 & -W & 2W \end{bmatrix}$$

and appears in the *i-th* diagonal block of size $mc \times mc$ $(1 \le i \le n)$; all other blocks are zero. Summing these block matrices over i leads to

$$\sum_{\substack{1\le i\le n \\ 1\le j\le m}} (\mathbf{D}M_{ij})^T W\, \mathbf{D}M_{ij} = Block\ Diag(\underbrace{\overline{W}_m^{\frac{1}{2}}\, \Lambda_m^T \Lambda_m\, \overline{W}_m^{\frac{1}{2}}\ , \cdots, \overline{W}_m^{\frac{1}{2}}\, \Lambda_m^T \Lambda_m\, \overline{W}_m^{\frac{1}{2}}}_{n-times}\ )$$

$$= \overline{W}_{nm}^{\frac{1}{2}} \cdot Block\ Diag(\underbrace{\Lambda_m^T \Lambda_m, \cdots, \Lambda_m^T \Lambda_m}_{n-times}) \cdot \overline{W}_{nm}^{\frac{1}{2}}$$

$$= \overline{W}_{nm}^{\frac{1}{2}} \cdot \Omega^T \Omega \cdot \overline{W}_{nm}^{\frac{1}{2}},$$

which establishes Equation 4.2.a.

Next, we'll show that $\mathbf{D}N_{nj}$ and $\mathbf{D}N_{ij}$ are block matrices of the following form:

$$\mathbf{D}N_{nj} = \left[0 \ \cdots \ 0 \ \ \mathrm{I} \ \ 0 \ \cdots \ 0 \,\vdots\, 0 \ \cdots \ 0 \,\vdots\, \cdots \,\vdots\, 0 \ \cdots \ 0 \,\vdots\, 0 \ \cdots \ 0 \ \ \text{-I} \ \ 0 \ \cdots \ 0\right]$$

$$(1 \le j \le m), \tag{4.3.5}$$

where I denotes the $c \times c$ identity matrix and appears in the $(1, j)$ - block, -I appears in the (n, j) - block, all other blocks are zero, and there are nm blocks of size $c \times c$. And

$$\mathbf{D}N_{ij} = \left[0 \ \cdots \ 0 \,\vdots\, 0 \ \cdots \ 0 \ \ \text{-I} \ \ 0 \ \cdots \ 0 \,\vdots\, 0 \ \cdots \ 0 \ \ \mathrm{I} \ \ 0 \ \cdots \ 0 \,\vdots\, 0 \ \cdots \ 0\right]$$

$$(1 \le i \le n-1\ ,\ 1 \le j \le m), \tag{4.3.6}$$

where -I appears in the (i, j) - block, I appears in the $(i+1, j)$ - block, all other blocks are zero, and there are nm blocks of size $c \times c$.

Choose $1 \le j \le m$. The *k-th* component of $N_{nj} : R^{nmc} \to R^c$ is

$$N_{nj,k}(u) \equiv u_{1,j,k} - u_{n,j,k} \qquad (1 \le k \le c).$$

Thus, the $[k,(r,s,t)]$ entry of $\mathbf{D}N_{nj}$, denoted by $(\mathbf{D}N_{nj})_{k,(r,s,t)}$, is

$$(\mathbf{D}N_{nj})_{k,(r,s,t)} \equiv \frac{\partial N_{nj,k}}{\partial u_{r,s,t}} = \begin{cases} 0 & \text{if } s \ne j \\ 0 & \text{if } t \ne k \\ 0 & \text{if } r \ne 1,n \\ 1 & \text{if } r = 1, s = j, t = k \\ -1 & \text{if } r = n, s = j, t = k \end{cases} \qquad (r,s,t) \in Ind, \ 1 \le k \le c. \quad (4.3.7)$$

The (r,s)-*th* $c \times c$ block of $\mathbf{D}N_{nj}$ is denoted by $(\mathbf{D}N_{nj})_{*,(r,s,*)}$ $(1 \le r \le n, \ \ 1 \le s \le m)$. Its (k,t) entry is $\dfrac{\partial N_{nj,k}}{\partial u_{r,s,t}}$ $(1 \le k \le c, \ \ 1 \le t \le c)$. By Equation 4.3.7, $(\mathbf{D}N_{nj})_{*,(r,s,*)}$ satisfies

$$(\mathbf{D}N_{nj})_{*,(r,s,*)} = \begin{cases} 0 & \text{if } s \ne j \\ 0 & \text{if } r \ne 1,n \\ \mathrm{I} & \text{if } r = 1, s = j \\ \text{-I} & \text{if } r = n, s = j \end{cases} \qquad (1 \le r \le n, \ \ 1 \le s \le m).$$

This establishes Equation 4.3.5.

Choose $1 \le j \le m$ and $i \in \{1, 2, \cdots, n-1\}$. The *k-th* component of $N_{ij} : R^{nmc} \to R^c$ is

$$N_{ij,k}(u) \equiv u_{i+1,j,k} - u_{i,j,k} \qquad (1 \le k \le c).$$

Thus, the $[k,(r,s,t)]$ entry of $\mathbf{D}N_{ij}$, denoted by $(\mathbf{D}N_{ij})_{k,(r,s,t)}$, is

$$(\mathbf{D}N_{ij})_{k,(r,s,t)} \equiv \frac{\partial N_{ij,k}}{\partial u_{r,s,t}} = \begin{cases} 0 & \text{if } s \neq j \\ 0 & \text{if } t \neq k \\ 0 & \text{if } r \neq i, i+1 \\ 1 & \text{if } r = i+1, s = j, t = k \\ -1 & \text{if } r = i, s = j, t = k \end{cases} \qquad (r,s,t) \in Ind, \ \ 1 \leq k \leq c. \quad (4.3.8)$$

The (r,s)-*th* $c \times c$ block of $\mathbf{D}N_{ij}$ is denoted by $(\mathbf{D}N_{ij})_{*,(r,s,*)}$ $(1 \leq r \leq n, \quad 1 \leq s \leq m)$. Its (k,t) entry is $\dfrac{\partial N_{ij,k}}{\partial u_{r,s,t}}$ $(1 \leq k \leq c, \quad 1 \leq t \leq c)$. By Equation 4.3.8, $(\mathbf{D}N_{ij})_{*,(r,s,*)}$ satisfies

$$(\mathbf{D}N_{ij})_{*,(r,s,*)} = \begin{cases} 0 & \text{if } s \neq j \\ 0 & \text{if } r \neq i, i+1 \\ \mathrm{I} & \text{if } r = i+1, s = j \\ \text{-}\mathrm{I} & \text{if } r = i, s = j \end{cases} \qquad (1 \leq r \leq n, \quad 1 \leq s \leq m).$$

This establishes Equation 4.3.6.

By Equation 4.3.5 the product $(\mathbf{D}N_{nj})^T W \, \mathbf{D}N_{nj}$ has the form

$$(\mathbf{D}N_{nj})^T W\,\mathbf{D}N_{nj} = \left[\begin{array}{ccccc|ccc|ccccc}
0 & & \cdots & & 0 & 0 & \cdots & 0 & 0 & & \cdots & & 0\\
& \ddots & & & & & & & & \ddots & & & \\
& & 0 & & & & & & & & 0 & & \\
& & W & & & & & & & & -W & & \\
& & & 0 & & & & & & & & 0 & \\
& & & & \ddots & & & & & & & & \ddots\\
0 & & \cdots & & 0 & 0 & \cdots & 0 & & & \cdots & & 0\\ \hline
0 & & \cdots & & 0 & 0 & & & 0 & & \cdots & & 0\\
& & & & & & \ddots & & & & & & \\
0 & & \cdots & & 0 & & & 0 & 0 & & \cdots & & 0\\ \hline
0 & & \cdots & & 0 & 0 & & 0 & 0 & & \cdots & & 0\\
& \ddots & & & & & & & & \ddots & & & \\
& & 0 & & & & & & & & 0 & & \\
& & -W & & & & & & & & W & & \\
& & & 0 & & & & & & & & 0 & \\
& & & & \ddots & & & & & & & & \ddots\\
0 & & \cdots & & 0 & 0 & & 0 & 0 & & \cdots & & 0
\end{array}\right]$$

$(1 \le j \le m)$,

where W appears in the $((1,j),(1,j))$ block and in the $((n,j),(n,j))$ block, $-W$ appears in the $((1,j),(n,j))$ block and in the $((n,j),(1,j))$ block, and all other blocks are zero. Consequently,

$$\sum_{j=1}^{m} (\mathbf{D}N_{nj})^T W\,\mathbf{D}N_{nj} = \left[\begin{array}{c|ccc|c}
\overline{W}_m & 0 & \cdots & 0 & -\overline{W}_m\\ \hline
0 & 0 & \cdots & 0 & 0\\
\vdots & & & & \vdots\\
0 & 0 & \cdots & 0 & 0\\ \hline
-\overline{W}_m & 0 & \cdots & 0 & \overline{W}_m
\end{array}\right],$$

where the size of the blocks is $mc \times mc$, $\overline{W}_m \equiv Block\ Diag(\underbrace{W, W, \cdots, W}_{m-times})$ appears in the $(1,1)$ block and in the (n,n) block, $-\overline{W}_m$ appears in the $(1,n)$ block and in the $(n,1)$ block, and all other blocks are zero.

By Equation 4.3.6, the product $(\mathbf{D}N_{ij})^T W\, \mathbf{D}N_{ij}$ has the form

$$
(\mathbf{D}N_{ij})^T W\, \mathbf{D}N_{ij} = \left[\begin{array}{ccc|ccccc|ccccc|ccc}
0 & \cdots & & & & & & & & & & & & & \cdots & 0 \\
\vdots & \ddots & & & & & & & & & & & & & & \vdots \\
 & & 0 & & & & & & & & & & & & & \\
\hline
 & & & 0 & & \cdots & & 0 & 0 & & \cdots & & 0 & & & \\
 & & & & \ddots & & & & & \ddots & & & & & & \\
 & & & & 0 & & & & & 0 & & & & & & \\
 & & & \vdots & & W & & \vdots & \vdots & & -W & & \vdots & & & \\
 & & & & & & 0 & & & & & 0 & & & & \\
 & & & & & & \ddots & & & & & \ddots & & & & \\
 & & & 0 & & \cdots & & 0 & 0 & & \cdots & & 0 & & & \\
\hline
 & & & 0 & & \cdots & & 0 & 0 & & \cdots & & 0 & & & \\
 & & & & \ddots & & & & & \ddots & & & & & & \\
 & & & & 0 & & & & & 0 & & & & & & \\
 & & & \vdots & & -W & & \vdots & \vdots & & W & & \vdots & & & \\
 & & & & & & 0 & & & & & 0 & & & & \\
 & & & & & & \ddots & & & & & \ddots & & & & \\
 & & & 0 & & \cdots & & 0 & 0 & & \cdots & & 0 & & & \\
\hline
 & & & & & & & & & & & & & 0 & & \\
\vdots & & & & & & & & & & & & & & \ddots & \vdots \\
0 & \cdots & & & & & & & & & & & & & \cdots & 0
\end{array}\right]
$$

$(1 \le i \le n-1,\ 1 \le j \le m)$,

where W appears in the $((i,j),(i,j))$ block and the $((i+1,j),(i+1,j))$ block, $-W$ appears in the $((i,j),(i+1,j))$ block and the $((i+1,j),(i,j))$ block, and all other blocks are zero. Consequently,

$$\sum_{j=1}^{m}(\mathbf{D}N_{ij})^T W\,\mathbf{D}N_{ij} = \left[\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & \overline{W}_m & -\overline{W}_m & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -\overline{W}_m & \overline{W}_m & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \end{array}\right] \qquad (1 \le i \le n-1),$$

where the size of the blocks is $mc \times mc$, $\overline{W}_m$ appears in the (i,i) block and in the $(i+1,i+1)$ block, $-\overline{W}_m$ appears in the $(i,i+1)$ block and in the $(i+1,i)$ block, and all other blocks are zero. Summing these block matrices over i leads to

$$\sum_{\substack{1\le i\le n \\ 1\le j\le m}}(\mathbf{D}N_{ij})^T W\,\mathbf{D}N_{ij} = \begin{bmatrix} 2\overline{W}_m & -\overline{W}_m & 0 & \cdots & 0 & -\overline{W}_m \\ -\overline{W}_m & 2\overline{W}_m & -\overline{W}_m & 0 & \cdots & 0 \\ 0 & -\overline{W}_m & 2\overline{W}_m & -\overline{W}_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\overline{W}_m & 2\overline{W}_m & -\overline{W}_m \\ -\overline{W}_m & 0 & \cdots & 0 & -\overline{W}_m & 2\overline{W}_m \end{bmatrix} = \overline{W}_{nm}^{\frac{1}{2}} \cdot \Lambda_n^T \Lambda_n \cdot \overline{W}_{nm}^{\frac{1}{2}},$$

which establishes Equation 4.3.b.

Proof of Proposition 4.4

By Propositions 4.2 and 4.3, $\mathbf{D}^2 F^+ = 2 \cdot \overline{W}_{nm} - \gamma \cdot \overline{W}_{nm}^{\frac{1}{2}} \cdot \left[\Omega^T\Omega + \Lambda_n^T\Lambda_n\right] \cdot \overline{W}_{nm}^{\frac{1}{2}}$.

Thus, as in the proof of Proposition 3.8,

$$u^T \mathbf{D}^2 F^+ u = 2 \cdot u^T \overline{W}_{nm} u - \gamma \cdot u^T \overline{W}_{nm}^{\frac{1}{2}} \cdot \left[\Omega^T \Omega + \Lambda_n^T \Lambda_n \right] \cdot \overline{W}_{nm}^{\frac{1}{2}} u$$

$$= 2 \cdot \left\| \overline{W}_{nm}^{\frac{1}{2}} u \right\|^2 - \gamma \cdot \left[\left\| \Omega \overline{W}_{nm}^{\frac{1}{2}} u \right\|^2 + \left\| \Lambda_n \overline{W}_{nm}^{\frac{1}{2}} u \right\|^2 \right]$$

$$\geq 2 \cdot \left\| \overline{W}_{nm}^{\frac{1}{2}} u \right\|^2 - \gamma \cdot \left[\| \Omega \|^2 + \| \Lambda_n \|^2 \right] \cdot \left\| \overline{W}_{nm}^{\frac{1}{2}} u \right\|^2$$

$$= \left[2 - \gamma \cdot \left(\| \Omega \|^2 + \| \Lambda_n \|^2 \right) \right] \cdot \left\| \overline{W}_{nm}^{\frac{1}{2}} u \right\|^2 .$$

Since $\gamma = 1/4$, $u^T \mathbf{D}^2 F^+ u \geq 0$ for all $u \in R^{nmc}$ provided $\| \Omega \|^2 + \| \Lambda_n \|^2 \leq 8$. But $\| \Omega \|^2$ and $\| \Lambda_n \|^2$ equal the largest eigenvalue of $\Omega^T \Omega$ and $\Lambda_n^T \Lambda_n$, respectively. These eigenvalues are bounded by 4 by Fact 3.8.1, Equation 3.3.8, and the fact that Λ_n has the same form as Λ and Ω is a block diagonal matrix with diagonal blocks equal to Λ_m, which has the same form as Λ. Therefore, $u^T \mathbf{D}^2 F^+ u \geq 0$ for all $u \in R^{nmc}$ and F^+ is convex.

5. REFERENCES

1. D. Mumford and J. Shah. "Boundary Detection by Minimizing Functionals," *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, Calif., 1985.

2. D. Mumford and J. Shah. "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems," *Commun. Pure Appl. Math.*, Vol. XLII, No. 4 (1989).

3. A. Blake and A. Zisserman. *Visual Reconstruction*, Boston, Mass., MIT Press, 1987.

4. D. Luenberger. *Linear and Nonlinear Programming*, 2nd ed., Reading, Mass, Addison-Wesley, 1984.

5. P. Lancaster. *Theory of Matrices*. New York and London, Academic Press, 1969.

INITIAL DISTRIBUTION

1 Naval War College, Newport (1E22, President)
1 Headquarters, 497 INOT, Falls Church (IOG Representative)
2 Defense Technical Information Center, Fort Belvoir
1 Center for Naval Analyses, Alexandria, VA (Technical Library)

ON-SITE DISTRIBUTION

3 Code 4TL000D (2 plus Archives copy)
21 Code 4T4100D
S. Chesnut (1)
J. Martin (20)